10M

УДК 004.89,004.93

К.В. Мурыгин
Институт проблем искусственного интеллекта МОН Украины и НАН Украины
г. Донецк, Украина
kir@iai.donetsk.ua

Особенности реализации алгоритма AdaBoost для обнаружения объектов на изображениях

В статье рассматриваются особенности реализации алгоритма AdaBoost для решения проблемы построения каскада классификаторов для эффективного обнаружения объектов на изображениях. В качестве элементарных классификаторов предлагается использовать прямоугольные логические свойства, что позволяет сделать независимым результат распознавания от монотонно возрастающих преобразований яркости изображений, не изменяющих их классовую принадлежность. Полученный в результате обучения каскад из 4 классификаторов позволяет решать задачу поиска на изображениях областей лиц в видеопотоке с параметрами 30 кадров размером 640×480 пикселей в секунду.

Введение

Алгоритм AdaBoost представляет собой эффективное средство обучения классификации. Особенностью используемого в нем подхода является принцип отбора и объединения набора простых и малоэффективных свойств в одно решающее правило, обладающее высокой классифицирующей способностью [1], [2]. Получение такого решающего правила является основной задачей обучения распознаванию образов. Во многих практических задачах получить его или его параметры не так просто, что связано со сложностью расположения объектов в пространстве признаков. С другой стороны, практически всегда можно эвристически сформировать целый набор достаточно простых характеристик [2], свойственных объектам распознавания, или предложить эвристическую процедуру для их генерации. Подобная процедура формирования свойств может, например, представлять собой объединение первичных признаков $x_1, \dots, x_n$ в виде произведения различных их степеней $y_k = \prod_{i=1}^{n} x_i^{\gamma_i^k}$, где индекс k определяет соответствующий набор $\{\gamma_1^k, \dots, \gamma_n^k\}$.

Такой способ формирования, вообще говоря, приводит к бесконечному множеству возможных свойств. Даже если каждое из них будет обладать слабой классифицирующей силой, то их совместное использование может дать вполне приемлемый по качеству результат. Алгоритм обучения AdaBoost строит решающее правило в виде линейной комбинации выходных значений слабых классификаторов, при этом достигается экспоненциальное уменьшение ошибки с увеличением числа свойств на обучающем наборе [3].

Если представить слабый (элементарный) классификатор двух классов Ω_1 и Ω_2, полученный на основе описанных выше простых свойств, в виде

$$h_k(\bar{x}) = \alpha y_k + \beta, \; h_k(\bar{x}) \in R,$$

классифицирующее свойство которого проявляется в смене знака на границе классов:

$$h_k(\bar{x}) = \begin{cases} > 0, \bar{x} \in \Omega_1, \\ \le 0, \bar{x} \in \Omega_2; \end{cases}$$

то в результате обучения будет получен классификатор вида

$$H(\bar{x}) = \sum_{t=1}^{T} a_t h_t(\bar{x}),$$

где T – число специальным образом отобранных простых классификаторов из общего их числа. Такой классификатор с учетом приведенного выше в качестве примера способа формирования простых свойств может представлять собой полином любой степени относительно исходных характеристик распознаваемых объектов (признаков) $x_1, \ldots, x_n$, что теоретически позволяет получить решающее правило любой сложности. При этом сложность определяется не эвристически, а генерируется в процессе обучения в зависимости от сложности межклассовой границы.

Процедура обучения по методу AdaBoost заключается в следующем [2]. Пусть заданы обучающие объекты, принадлежащие двум разным классам, в виде $(\vec{x}_1, y_1), \ldots, (\vec{x}_m, y_m)$, где $\bar{x}_i \in X$ – вектор признаков, определяющий объект распознавания, $y_i \in Y = \{-1, +1\}$ – число, определяющее его классовую принадлежность, m – объем обучающей выборки. Изначально каждому обучающему объекту задается одинаковый вес $D_1(i) = \frac{1}{m}$. После этого начинается циклический процесс обучения, который схематически может быть представлен следующим образом.

Для каждого цикла обучения $t = 1, \ldots, T$:

1) выбирается лучший элементарный классификатор $h_t(\bar{x}): X \to R$, дающий минимальную ошибку на обучающем наборе с учетом текущих весов объектов D_t. В более простом случае используются классификаторы вида $h_t(\bar{x}): X \to \{-1; +1\}$;

2) выбирается действительное число $\alpha_t > 0$;

3) выполняется перевзвешивание обучающего набора согласно следующей зависимости $D_{t+1}(i) = \frac{D_t(i)\exp(-\alpha_t y_i h_t(\bar{x}_i))}{Z_t}$, где Z_t – нормировочный множитель, такой, что сумма весов всех обучающих объектов должна равняться 1.

Таким образом, в ходе обучения формируется набор классификаторов $h_t(\bar{x})$ и чисел α_t. Итоговое решающее правило, комбинирующее найденные элементарные классификаторы $h_t(\bar{x})$, представляет собой их линейную комбинацию:

$$H(\bar{x}) = \operatorname{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(\bar{x})\right).$$

Приведенный процесс продолжается до достижения необходимой величины ошибки классификации. Как видно из выражения для получения новых весов обучающих объектов $D_{t+1}(i)$, в ходе перевзвешивания веса правильно распознанных на текущем цикле обучения объектов уменьшаются, неправильно – увеличиваются. Таким образом, каждый следующий цикл обучения будет концентрироваться на

плохо разделенных предыдущими циклами объектах, постепенно уточняя итоговое решающее правило.

В работах [1], [2] показано, что для оптимизации процесса обучения с точки зрения максимизации скорости уменьшения ошибки в зависимости от цикла обучения, значение параметра α_t, используемого при перевзвешивании обучающего набора, должно выбираться согласно зависимости $\alpha_t = \frac{1}{2}\ln\left(\frac{1-e_t}{e_t}\right)$, где e_t – ошибка классификации на t-м цикле обучения, полученная с учетом весов обучающих объектов D_t.

Алгоритм обучения AdaBoost может быть успешно применен для решения задач распознавания изображений, в частности для быстрого поиска на изображении произвольных объектов [4]. Требования к быстродействию алгоритмов классификации при решении последней задачи являются очень жесткими [5]. Для обеспечения требуемого быстродействия в работе [4] предложены два основных принципа.

1. Использовать в качестве элементарных классификаторов $h_t(\bar{x})$ прямоугольные, просто масштабируемые свойства, определить значения которых можно с достаточно малыми вычислительными затратами по интегральному изображению.

2. Использовать не один классификатор, а целый набор классификаторов, объединенных в каскад, что существенно увеличивает среднюю скорость проведения классификации, так как на следующий этап обработки поступают только области исходного изображения, распознанные предыдущими классификаторами как искомый объект.

Предложенные в работе [4] свойства являются ограниченным подмножеством возможных прямоугольных свойств. Ряд свойств, предложенных в [6], требуют дополнительных усилий для получения их значений. В частности, для получения средней яркости в повернутых прямоугольных областях необходим предварительный расчет интегрального изображения для повернутого на 45 градусов исходного изображения. Кроме этого, предложенный способ построения классификатора на их основе не решает в полной мере проблему зависимости значения классификатора от условий получения изображения. Данная статья посвящена исследованиям вопроса выбора более эффективного с точки зрения вычислительной сложности набора элементарных классификаторов $h_t(\bar{x})$, основанных на использовании прямоугольных свойств [4] более широкого вида, и сравнительной оценке их эффективности на примере решения задачи обнаружения лиц на изображении.

Использование Хаар-подобных свойств для описания свойств изображений

Основными критериями качества признака для решения широкого круга задач, и в особенности задач распознавания зрительных образов, являются его разделительные свойства и сложность его получения, учитывая необходимость быстрого поиска области объекта, что предполагает классификацию большого числа кандидатов при обработке одного изображения [5].

С точки зрения необходимости использования достаточно простых алгоритмов получения признаков, перспективным является использование Хаар-подобных свойств, в отличие от предложенных в [4] и [6], представляющих собой результат сравнения яркостей в двух прямоугольных областях изображения (рис. 1).

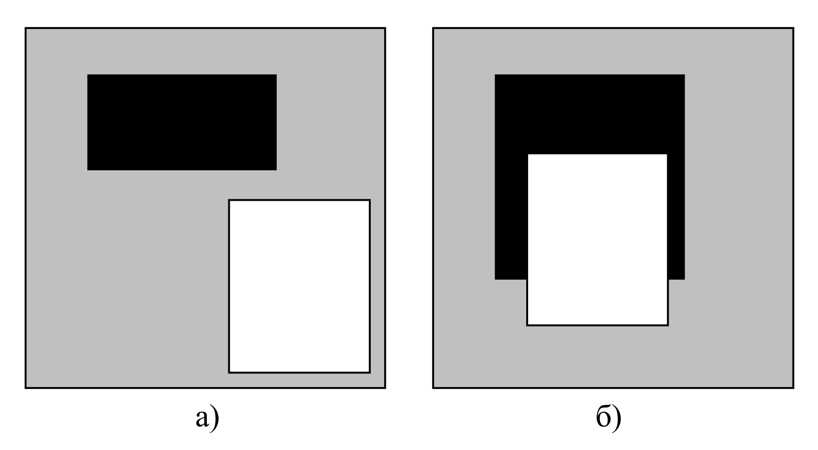


Рисунок 1 – Вид прямоугольных свойств, используемых в качестве признаков при обнаружении объектов: а) – области не пересекаются; б) – области пересекаются

Значение признака для данной области изображения или отклик области изображения на данное свойство вычисляется на основе следующего выражения:

$$R = \frac{S_Б}{N_Б} - \frac{S_Ч}{N_Ч} \quad (1)$$

в случае непересекающихся областей (рис. 1а), и

$$R = \frac{S_Б}{N_Б} - \frac{S_Ч - S_{Ч \cap Б}}{N_Ч - N_{Ч \cap Б}} \quad (2)$$

в случае пересечения областей (рис. 1б). Здесь индексы Ч и Б означают черную и белую области соответственно, а $Ч \cap Б$ обозначает область пересечения областей черного и белого цвета; S – сумма яркостей пикселей изображения, находящихся под областью; N – число пикселей изображения, находящихся под областью. Таким образом, соответствующие отклики означают разность средних яркостей пикселей реальной области изображения, находящейся под белой частью изображения свойства, и находящихся под его черной частью (рис. 1а, б). Отклик такого свойства не зависит от масштаба изображения и смещения изображения по шкале яркости.

Кроме приведенных выше выражений для отклика признака на конкретную область изображения (1) и (2) в качестве значения отклика на свойство можно использовать и следующие выражения:

1) в случае непересекающихся областей

$$R = S_Б - S_Ч, \quad (3)$$

В случае пересечения областей

$$R = S_Б - \left(S_Ч - S_{Ч \cap Б}\right); \quad (4)$$

2) в случае непересекающихся областей

$$R = \begin{cases} 1, \dfrac{S_Б}{N_Б} > \dfrac{S_Ч}{N_Ч}; \\ -1, \dfrac{S_Б}{N_Б} \le \dfrac{S_Ч}{N_Ч}. \end{cases} \quad (5)$$

В случае пересечения областей

$$R = \begin{cases} 1, \dfrac{S_Б}{N_Б} > \dfrac{S_Ч - S_{Ч \cap Б}}{N_Ч - N_{Ч \cap Б}}; \\ -1, \dfrac{S_Б}{N_Б} \le \dfrac{S_Ч - S_{Ч \cap Б}}{N_Ч - N_{Ч \cap Б}}. \end{cases} \quad (6)$$

С точки зрения инвариантности значения свойства относительно возможных яркостных трансформаций изображения использование для получения значения отклика выражений (5) и (6) является наиболее приемлемым. Значения, получаемые на основе этих выражений, являются инвариантными по отношению к любым линейным поэлементным монотонно возрастающим яркостным преобразованиям изображений, которые не изменяют его классовую принадлежность, но могут существенно изменять распределения яркостей в изображении. К таким преобразованиям относятся контрастирование и изменение яркости изображения, которые могут автоматически применятся при захвате видеоданных в разнообразных устройствах получения видеоизображений для их корректировки и улучшения качества.

Обучение классификации

Исследования эффективности предложенных свойств для решения задач обнаружения объектов на изображениях проводились на задаче обнаружения лиц. Обучение классификации лицо/фон выполнялось на основе метода AdaBoost. Для обучения использовалась база данных изображений лиц, содержащая 2000 изображений. Каждое изображение было размечено вручную и масштабировано к масштабу 12×12 пикселей. В базе данных содержались изображения лиц людей разных рас, полученные при разном угле и интенсивности освещения, изображения лиц с очками различной формы, бородой и (или) усами, имеющие отклонения от фронтального ракурса съемки до 30 градусов. Для расширения базы данных и учета изменения направления освещения она была дополнена зеркальными отражениями каждого изображения лица. Таким образом, итоговая база изображений лиц содержала 4000 изображений размером 12×12 пикселей.

Для формирования базы изображений фона была разработана специальная программа, которая на основе сформированного набора изображений разного размера (около 400 изображений), не содержащих лица, случайным образом формировала 100 000 изображений. При этом случайным образом выбирались как положения области фона на изображения, так и ее масштаб. Предполагая необходимость формирования в конечном итоге каскада классификаторов, в программе была учтена возможность формирования набора изображений фона, удовлетворяющих заданному условию, в частности условию прохождения изображения фона через все предыдущие классификаторы. Таким образом, программа позволяет формировать базу изображений фона, трактуемых уже обученными классификаторами как лица. Необходимость подобной программы объясняется тем фактом, что после обучения ошибка ложного обнаружения должна быть очень низкой (порядка $<10^{-6}$), а соответственно число обучающих объектов фона как минимум больше 10^6, что при использовании алгоритма обучения AdaBoost влечет значительные вычислительные затраты, существенно затрудняющие анализ получаемых результатов. Необходимость достижения такой низкой ошибки ложного обнаружения объясняется значительным числом кандидатов лиц на обрабатываемом изображении, а предполагаемый процесс поиска основывается на классификации всех возможных кандидатов с учетом выбранного шага по масштабам.

Процесс поиска наиболее эффективного свойства на каждом этапе обучения, предусмотренном используемым алгоритмом обучения, предполагает перебор всех возможных прямоугольных свойств, и выбор из них наилучшего, в смысле минимума ошибки, вычисленной по взвешенной обучающей базе данных изображений, соответствующей данному этапу обучения. Так как число вариантов прямоугольных свойств даже для используемого масштаба изображений велико, чтобы осуществить полный перебор (порядка 4×10^7, то есть для полного перебора необходимо выполнить указанное число оценок ошибок классификации по имеющейся базе данных обучающих объектов, а это сама по себе вычислительно сложная задача, учитывая размеры обучающих выборок), поиск наилучших свойств выполнялся в два этапа. На первом этапе выполнялся поиск свойства, дающего наименьшую ошибку из всех отобранных случайным образом. На втором этапе методом последовательных приближений найденное на первом этапе прямоугольное свойство уточнялось по критерию минимизации ошибки. Такой подход позволил успешно проводить обучение каждого классификатора каскада за удовлетворительное время (0,5 – 3 часа).

В ходе обучения необходимо учитывать необходимость включения в исследуемый набор свойств и классификаторов, которые с точки зрения классификации взвешенного обучающего набора не имеют смысла, например, классифицируют все изображения как 1 объект. Если таких свойств не будет, то скорость спада ошибки обучения в зависимости от цикла обучения может существенно замедлиться. Это связано с балансом весов объектов одного и другого классов. В ходе обучения этот баланс может существенно смещаться как в одну, так и в другую сторону и ошибка принятия всех объектов 1-го класса за 2-й класс может оказаться наименьшей из всех других исследованных классификаторов на данном этапе. Введение подобных «бесполезных» классификаторов позволяет поддерживать при обучении приемлемый баланс между суммарными весами объектов обоих классов. С точки зрения итогового решающего правила в виде линейной комбинации элементарных классификаторов это означает смещение классифицирующей поверхности, определяемой линейной комбинацией простых классификаторов, относительно начала координат.

Для ускорения процесса обработки изображений с целью обнаружения лиц, как отмечалось в [5], количество свойств, используемых на начальных каскадах, должно быть, с одной стороны, минимально возможным, а с другой, обеспечивать допустимый пропуск ложных лиц на следующие, более вычислительно сложные и обеспечивающие более тонкое разделение лиц и фона, этапы обработки.

В ходе обучения каскадов классификаторов согласно методу AdaBoost для прямоугольных свойств, задаваемых выражениями (5) и (6), для первого каскада была получена следующая зависимость ошибки ложного обнаружения от числа используемых признаков (рис. 2).

Ограничение на ошибку пропуска лиц – не более 1% – принято, так как для решения задачи классификации с заданным качеством (95% правильного обнаружения) планировалось использовать не более 5 классификаторов в каскаде.

Как видно из приведенной на рис. 2 зависимости, значения ошибки ложной тревоги достаточно быстро уменьшаются с увеличением числа используемых прямоугольных свойств. При этом функцию ошибки можно с удовлетворительной точностью в области больших значений числа используемых свойств аппроксимировать экспоненциальной зависимостью. Рассчитанные параметры и вид этой аппроксимирующей функции приведен на рис. 2. Из графика также видно, что для отсеивания на первом этапе 99% ложных лиц достаточно использовать 47 прямоугольных признаков и линейную разделяющую функцию, получаемую на выходе алгоритма обучения.

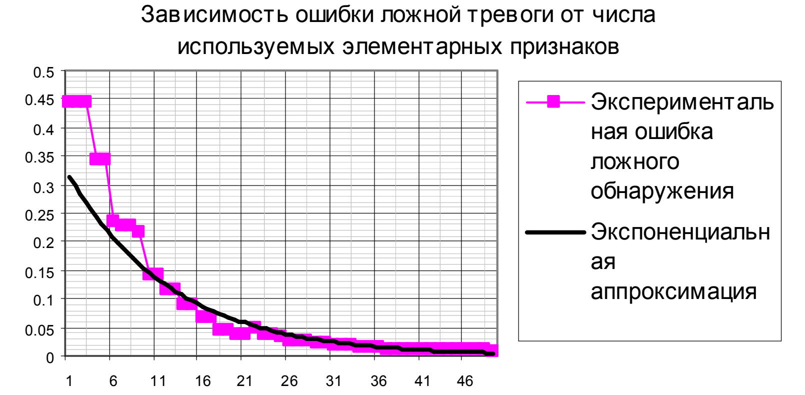


Рисунок 2 – Зависимость ошибки ложной тревоги от числа используемых элементарных признаков, при ограничении на ошибку пропуска лица 1%

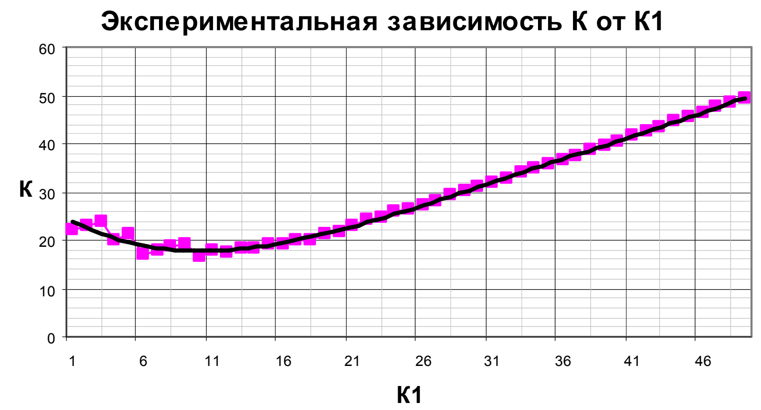


Рисунок 3 – Зависимость среднего числа используемых при прохождении двух каскадов прямоугольных свойств (К) от числа свойств, используемых на промежуточном первом каскаде (К1)

Приведенная зависимость позволяет оценить вычислительную сложность алгоритма классификации, состоящего из двух каскадов. Если мы зададимся ограничением ошибки пропуска лиц в 2% и требованием к ошибке ложного обнаружения – 1%, то для среднего числа используемых свойств при прохождении двух каскадов будем иметь зависимость, приведенную на рис. 3. При получении этой зависимости учитывалось, что второй классификатор каскада содержит 47 свойств, что гарантированно обеспечивает ошибку ложного обнаружения менее 1%.

Анализируя приведенные данные можно заключить, что в данном случае оптимальным с точки зрения минимизации вычислительной сложности алгоритма при прохождении каскада из двух классификаторов является использование в первом классификаторе 10 прямоугольных признаков. Это, согласно приведенным на рис. 3 данным, позволяет уменьшить вычислительные затраты более чем 2,5 раза по сравнению с использованием только одного второго классификатора. Получение подобных зависимостей позволяет управлять вычислительной сложностью алгоритма каскадной классификации и настраивать этот параметр исходя из требований к скорости обнаружения лица на изображении.

Наряду с исследованием скоростных характеристик алгоритма каскадной классификации были проведены исследования по определению зависимости качества классификаторов от используемых выражений для определения отклика прямоугольных свойств на изображение. Методом AdaBoost были экспериментально получены зависимости ошибки

классификации от числа используемых свойств в случае использования выражений для отклика изображения на прямоугольное свойство (1, 2) и (5, 6). Сравнительные данные приведены на рис. 4.

Исходя из приведенных данных можно говорить о том, что при одном количестве используемых прямоугольных свойств суммарная ошибка классификации лицо/фон при использовании выражений (1, 2) меньше, чем при использовании выражений (5, 6). Используя отклики прямоугольных свойств, получаемые на основе выражений (1, 2), был получен каскад из 4 классификаторов, содержащих соответственно 5, 8, 100, 200 признаков. При этом два первых классификатора использовали квадратическую разделяющую функцию, полученную методом, описанным в [7], а два последних – линейную, полученную методом AdaBoost. Достигнутые после обучения классификаторов показатели качества составили 0,04 – вероятность пропуска лиц, и 8×10^{-6} – вероятность ложного обнаружения. При этом была получена скорость обнаружения – 30 кадров размером 640×480 пикселей в секунду при поиске лиц в диапазоне $\left[{}^{H}\!/_{6}; {}^{2H}\!/_{3}\right]$ с шагом в 1,1 раза на компьютере Pentium Core2Duo с частотой процессора 2,33 ГГц. Здесь H – вертикальный размер изображения.

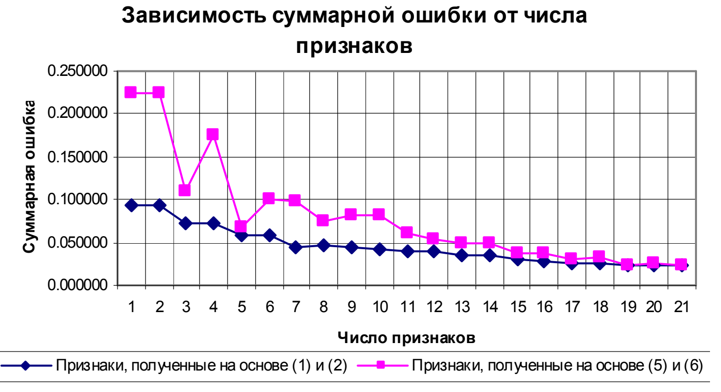


Рисунок 4 – Зависимость суммарной ошибки от числа используемых признаков для разных способов получения отклика свойства на изображение

В результате дополнительного цикла обучения был получен также каскад из 4 классификаторов, использующих выражения для откликов на прямоугольные свойства (5, 6). В ходе обучения были достигнуты такие же показатели по качеству обнаружения, что и при использовании выражений (1, 2) при увеличении вычислительной сложности приблизительно в 2 раза.

Выводы

Рассмотренный в статье подход к решению задачи обнаружения объектов на изображении на основе использования алгоритма обучения AdaBoost показал удовлетворительные результаты в применении его к задаче классификации лицо/фон. Полученный в результате обучения каскад из 4 классификаторов позволяет использовать его для обнаружения областей лиц в кадрах видеопоследовательности, при этом достигнута скорость обработки 30 кадров в секунду для размеров входных изображений 640×480 пикселей. Приведенная вычислительная производительность алгоритма получена при поиске лиц в диапазоне $\left[{}^{H}\!/_{6}; {}^{2H}\!/_{3}\right]$ с шагом в 1,1 раза на компьютере Pentium Core2Duo с частотой процессора 2,33 ГГц (здесь H – вертикальный размер изображения). Достиг-

нутые после обучения классификаторов показатели качества составили 0,04 – вероятность пропуска лиц, и 8×10^{-6} – вероятность ложного обнаружения. Основным направлением дальнейших исследований является исследование возможностей использования описанного подхода для обнаружения других целевых объектов, в частности деталей лица, для определения степени его инвариантности к свойствам объектов поиска. Кроме этого, рассмотренный метод позволяет модифицировать способ представления классификаторов в направлении рассмотрения вершин многомерного куба как дескрипторов распознаваемых классов. Такой способ представления позволит существенно сократить вычислительные затраты при условии древовидного представления положения объектов в вершинах.

Литература

1. Schapire R.E. Improved Boosting Algorithms Using Confidence-rated Predictions / R.E. Schapire, Y.Singer // Proceedings of Machine Learning. – 1999. – P. 297-336.
2. Schapire. Robert E. The boosting approach to machine learning: An overview / Robert.E. Schapire // In MSRI Workshop on Nonlinear Estimation and Classification. – 2002.
3. Schapire. Robert E. Boosting the margin: A new explanation for the effectiveness of voting methods / Robert E. Schapire. Freund Y. Bartlett P, Lee W.S. // Proc. Of the Fourteenth International Conference on Machine Learning. – 1997.
4. Paul Viola. Robust real-time object detection / Paul Viola and Michael J. Jones. // Proc. of IEEE Workshop on Statistical and Computational Theories of Vision. – 2001.
5. Мурыгин К.В. Обнаружение объектов на изображении на основе каскада классификаторов / К.В. Мурыгин // Искусственный интеллект. – 2007. – № 2. – С. 104-108.
6. Bradski G. Learning-based computer vision with Intel's open source computer vision library [Электронный ресурс] / G. Bradski, A. Kaehler, V. Pisarevsky // Intel Technology Journal. – Режим доступа : http://develiper.intel.com/technology/itj-/index.htm.
7. Мурыгин К.В. Построение классификаторов на основе разделяющих поверхностей / К.В. Мурыгин // Искусственный интеллект. – 2008. – № 2. – С. 65-69.

К.В. Муригін

Особливості реалізації алгоритму AdaBoost для виявлення об'єктів на зображеннях

У статті розглядаються особливості реалізації алгоритму AdaBoost для вирішення проблеми побудови каскаду класифікаторів для ефективного виявлення об'єктів на зображеннях. Як елементарні класифікатори пропонується використовувати прямокутні логічні властивості, що дозволяє зробити незалежним результат розпізнавання від монотонно зростаючих перетворень яскравості зображень, що не змінюють їх класову приналежність. Отриманий у результаті навчання каскад з 4 класифікаторів дозволяє вирішувати задачу пошуку на зображеннях областей облич у відеопотоці з параметрами 30 кадрів розміром 640×480 пікселів за секунду.

K.V. Murygin

The Features of Algorithm AdaBoost Implementation for Objects Detection on the Images

In article features of implementation of AdaBoost algorithm for solution a problem of construction of the classifiers cascade for effective detection of objects on images are considered. As elementary classifiers it is offered to use rectangular logical properties that allow to make independent a result of recognition from monotonously increasing conversions of brightness of the images which don't change their class association. The cascade received as a result of training process 4 classifiers which allows to solve the search task of face areas in a videostream with parameters of 30 frames of the size 640×480 pixels per second.

Статья поступила в редакцию 11.06.2009.